

# 逐步判别分析在侵蚀土壤分类中的应用

杨艳生 史德明

(中国科学院南京土壤研究所)

侵蚀土壤分类是研究土壤侵蚀、开展水土保持的重要内容。不同侵蚀土壤类型反映了土壤侵蚀的历史面貌和当前土壤肥力状况。研究侵蚀土壤分类,可为土壤的合理利用、改良和侵蚀防治提供依据。

过去,侵蚀土壤类型划分是以当前所保留的土壤发生层厚度为依据。保留的土壤发生层厚度是土壤诸侵蚀因子综合作用的结果,因此,需要研究各侵蚀因子与侵蚀土壤类型的关系,找出导致侵蚀土壤类型形成的主要因子,并根据各因子强度来划分侵蚀土壤类型,使侵蚀土壤分类数值化。但是,进行数值分类,需要应用多元分析方法,并借助电子计算机进行数据处理。本文以宁夏固原县的工作为例,说明逐步判别分析在侵蚀土壤分类中的应用。

## 一、判别分析的一般原理

判别分析是一种多元数值分析方法[1,2],用以土壤分类时,是根据各样品的观测数据,找出决定这些样品类型归属的判别函数,据此对样品进行分类。在进行数据处理时,通常把所观测的土壤剖面作为样品,所进行观测的各侵蚀因子强度作为变量。若现有两组样品,如果只有一个变量,并且各组样品的变量观测值相差悬殊或很少重迭时,这两组样品就易于分开;但如果两组样品的变量观测值相重迭,单用一个变量就难以将这两组样品分开,这时需要考虑增加变量数目,以达到样品分组的目的。变量增加以后,重迭现象还可能出现,进行样品分组仍有困难,判别分析就是解决这一难题的一种方法。应用判别分析就是要在两组样品间找到能合理地区分它们的判别线(或平面),并求出相应的判别函数,并利用这个判别函数,对样品进行类型判别。

在分类时,当然不只限于两组样品。对于多组样品则有相应的多组判别分析法。多组判别就是要求在多维空间,以某种规则划分出若干互不相交的区域,使样品分别归于若干组,并要求这一分组达到最合理。假定现有三组样品和两个变量,就要求在二维空间求出

三根线,划分出  $R_1$ 、 $R_2$ 、 $R_3$  三个子空间,使分属于这三个子空间的三组样品分组误差最小。所以判别分析的基本思想就是把同一组样品看作是有  $V$  个变量组成的一个总体,这个总体在  $V$  维空间形成一个点集,其它样品总体,也可按同样的方式去描述。判别函数就是经过计算确定下来,能有效地将这些点集合理分开的  $(V-1)$  维平面函数。

在进行判别分组时,如果变量数太少,分组的效果可能不好,因此需要增加变量以提高判别效果。但是前人已有证明[1],当变量数增加到一定时,再增加变量数,反而会使判别效果降低,而且变量数的增加,势必使观测和计算的工作量也增加。逐步判别分析就是以选取最少变量数达到最佳判别效果的一种分析方法。本文介绍费歇尔准则下的逐步判别分析的应用,通过分析(1)选出能判定侵蚀土壤类型归属的变量数;(2)对原有的土壤类型进行判别调整;(3)建立判定土壤类型归属的判别函数。

本文的计算系参照文献[1],并在七〇九机上实现计算。

## 二、侵蚀土壤的类型指标

侵蚀土壤类型是过去和当前各侵蚀因子综合影响的结果。因此,可以通过研究侵蚀土壤类型和侵蚀因子间的关系,采用侵蚀因子的强度指标来判定土壤类型归属。这样,主要侵蚀因子的强度指标,就成为侵蚀土类划分的依据,并通过计算使侵蚀土壤分类数值化。

1. 侵蚀土壤的野外分类指标 在野外,不同侵蚀土类,主要是通过观测土壤保留的发生层厚度,和观察到的一些土壤侵蚀因子的综合概念来划分(表1)。这一划分方法,指标较为单纯明确,并基本能反映侵蚀土壤的实际面貌。但对过渡性土类,缺乏明确的划分标准,没有反映侵蚀土类与主要侵蚀因子的数值关系。

2. 主要侵蚀因子及其强度分级 为了对侵蚀土

表 1

侵 蚀 土 壤 野 外 分 类 指 标

侵 蚀 土 类	代 号	主 要 划 分 指 标	
		保 留 土 壤 发 生 层	其 它
轻 微 侵 蚀 土 壤	1	A层几乎没有被侵蚀, B、C层完整	地面覆盖度100%
轻 度 侵 蚀 土 壤	2	A层遭不同程度侵蚀, B、C层完整	地面较平, 覆盖度大于80%
中 度 侵 蚀 土 壤	3	无A层, B层几乎未被侵蚀	坡度中等, 覆盖度50—80%
强 度 侵 蚀 土 壤	4	无A层, B层遭不同程度侵蚀	坡度较陡, 覆盖度30—50%
极 强 度 侵 蚀 土 壤	5	无B层, A层极浅薄或C层层出露	坡度陡, 覆盖度小于30%

壤进行判别分类, 必须首先确定主要侵蚀因子, 并进行强度分级。主要侵蚀因子就是对土壤侵蚀起主要作用的因子, 一般包括降雨、地形、土壤特性和人为活动等。准确地确定主要侵蚀因子, 并进行确切的强度分级是取得判别分类良好效果的关键。有些主要侵蚀因子, 常常因地而异, 需经过仔细的调查分析才能明

确。本文对各观测土壤的主要侵蚀因子强度指标均采用数码记值[3,4]。各数码值大小与强度级别的高低一致。各侵蚀因子强度分级如表 2。各观测土壤的侵蚀因子强度值如表 3。侵蚀土壤分类的数值化, 就是以表 3 的数据资料, 通过逐步判别分析处理后实现的。

表 2

各 侵 蚀 因 子 强 度 分 级

因 子 代 号	主 要 侵 蚀 因 子	强 度 分 级				
		1	2	3	4	5
1	土 壤 可 蚀 性	可 蚀 性 K 值 *	小	较 小	较 大	大
		数 码 取 值	1	5	8	10
2	地 面 坡 度	地 面 坡 度 (度)	0—5	5—10	10—15	15—25
		数 码 取 值	1	3	6	10
3	土 壤 利 用 状 况	土 壤 利 用 方 式 或 小 型 地 貌 单 元 类 型	林 地 或 川 地	草 地 或 台 地	荒 地 或 缓 坡 农 地	坡 耕 地 或 光 坡 地
		数 码 取 值	1	5	10	20
4	土 壤 利 用 干 扰 强 度	人 类 不 合 理 利 用 对 土 壤 干 扰 的 程 度	轻	中	强	很 强
		数 码 取 值	1	5	10	15

\*可 蚀 性 K 值 参 见 文 献 [5]。

### 三、判别分析计算的主要结果及应用

通过分析计算, 现将组内离差矩阵W、总离差矩阵T列于表 4; 各步计算相应选取的变量和检验值列于表 5; 最后得出的判别系数列于表 6。

利用表 6 判别系数, 即可建立起判别方程, 判别方程的一般形式是:

$$F_i = \ln Q_i + \sum_{k=1}^m x_k c_{ki} + c_{0i}$$

其中: i 为侵蚀土壤类型代号。侵蚀土壤共分五类, i 的取值从 1 至 5;

m 为通过计算选取的变量个数。本计算四项变

量全部选取。故 m 的取值为 4;

Q<sub>i</sub> 为各类土壤的先验概率, 可近似用此类土壤观

表 4 组内离差矩阵W、总离差矩阵T

		W 阵			
		532.53	-351.66	312.21	-26.77
T 阵			926.81	-163.43	-277.86
		840.32		1978.76	-542.69
		-109.07	1769.57		1122.78
		997.04	998.21	4187.56	
		407.54	641.21	1097.72	2413.56

表3 各观测土壤侵蚀因子强度值及计算分类

* 算前 分类 号	* 计算 分类 号	侵蚀因子强度值				土壤 序 号	* 算前 分类 号	* 计算 分类 号	侵蚀因子强度值				土壤 序 号
		土壤 可蚀性	地面坡度	土壤利用 状况	利用干扰 强度				土壤 可蚀性	地面坡度	土壤利用 状况	利用干扰 强度	
1	1	5	1	1	1	1	3	4	5	10	5	10	43
1	1	5	1	1	1	2	3	3	5	6	10	5	44
1	1	5	1	1	1	3	4	4	5	10	20	1	45
1	1	5	1	1	1	4	4	4	5	10	20	1	46
1	2	1	6	5	1	5	4	4	1	15	10	5	47
1	1	1	3	5	1	6	4	2	1	15	1	1	48
1	1	1	1	5	1	7	4	2	1	10	1	5	49
1	1	5	1	1	1	8	4	4	1	15	5	10	50
1	1	5	1	1	1	9	4	4	5	10	10	15	51
1	1	1	3	1	5	10	4	4	5	10	20	5	52
2	1	5	1	1	1	11	4	4	5	6	10	15	53
2	1	8	1	1	1	12	4	3	1	19	1	10	54
2	1	5	1	1	1	13	4	4	5	6	10	15	55
2	1	5	1	1	1	14	4	4	5	6	10	15	56
2	2	1	10	1	1	15	4	4	5	15	20	1	57
2	2	1	15	1	1	16	4	4	5	6	10	15	58
2	2	1	10	1	5	17	4	4	5	6	20	10	59
2	2	1	6	1	5	18	4	4	5	15	5	10	60
2	3	5	1	20	1	19	4	4	5	10	10	10	61
2	3	1	10	10	5	20	4	4	1	10	5	15	62
2	3	1	10	5	5	21	4	4	1	15	10	10	63
2	2	8	1	5	1	22	4	4	1	15	10	10	64
2	2	5	3	10	1	23	5	5	10	10	20	10	65
2	2	8	1	1	5	24	5	5	10	10	20	10	66
3	3	5	6	10	5	25	5	5	10	15	20	5	67
3	3	5	3	10	10	26	5	5	10	6	20	15	68
3	3	1	10	10	5	27	5	5	10	15	20	5	69
3	3	5	6	10	5	28	5	5	10	6	20	15	70
3	3	5	3	5	15	29	5	5	10	10	20	15	71
3	3	1	10	10	5	30	5	5	10	15	20	10	72
3	3	1	6	10	10	31	5	4	1	15	20	10	73
3	3	1	10	10	5	32	5	4	1	10	10	15	74
3	3	5	6	5	10	33	5	4	1	10	10	15	75
3	3	5	3	5	15	34	5	4	1	15	10	15	76
3	3	5	6	10	10	35	5	5	10	10	20	15	77
3	3	5	10	10	5	36	5	5	10	6	20	15	78
3	3	5	10	10	5	37	5	5	10	10	10	15	79
3	3	5	3	10	15	38	5	5	10	6	20	10	80
3	4	5	6	5	15	39	5	5	10	6	20	10	81
3	3	5	3	20	5	40	5	5	10	6	10	15	82
3	3	5	3	20	5	41	5	5	10	15	20	15	83
3	4	5	10	5	10	42	5	5	10	15	20	15	84

\* 代号表示的土类同表1。

表5 各步计算选取的变量数和检验值

计算步数	选取的变量*	F统计量	累计引进的变量数	U统计量	$\chi^2$ 统计量
1	$x_4$	22.7	1	0.47	61.22
2	$x_3$	22.4	2	0.21	121.65
3	$x_2$	15.6	3	0.12	167.99
4	$x_1$	7.0	4	0.09	191.67

\* 变量的足码同表2侵蚀因子代号

表6 各类土壤的判别系数

土类代号	$c_0$	$c_4$	$c_3$	$c_2$	$c_1$
1	-2.41	0.31	0.09	0.61	0.87
2	-6.45	0.59	0.23	1.12	1.21
3	-17.43	1.36	0.67	1.63	1.38
4	-25.37	1.57	0.78	2.13	1.53
5	-47.91	2.15	1.13	2.69	2.45

测数与观测的土壤总数之比代替。

由于观测的土壤总数为84, 各类土壤的先验概率和相应 $\ln Q_i$ 值如下:

$$\begin{aligned} Q_1 &= \frac{10}{84}, & \ln Q_1 &= -2.13 \\ Q_2 &= \frac{14}{84}, & \ln Q_2 &= -1.79 \\ Q_3 &= \frac{20}{84}, & \ln Q_3 &= -1.44 \\ Q_4 &= \frac{20}{84}, & \ln Q_4 &= -1.44 \\ Q_5 &= \frac{20}{84}, & \ln Q_5 &= -1.44 \end{aligned}$$

所得到的一组判别方程为:

$$F_1 = -4.54 + 0.31x_4 + 0.09x_3 + 0.61x_2 + 0.87x_1$$

$$F_2 = -8.24 + 0.59x_4 + 0.23x_3 + 1.12x_2 + 1.21x_1$$

$$F_3 = -18.87 + 1.36x_4 + 0.67x_3 + 1.63x_2 + 1.38x_1$$

$$F_4 = -26.81 + 1.57x_4 + 0.78x_3 + 2.13x_2 + 1.53x_1$$

$$F_5 = -49.35 + 2.15x_4 + 1.13x_3 + 2.69x_2 + 2.45x_1$$

方程中F和x的足码分别是表1中的土类代号和表2中的侵蚀因子代号。这一判别方程组有如下两方面的实际应用。

1. 判定侵蚀土壤的类型 首先利用该方程组可对原有的侵蚀土类进行判定, 得出的计算分类结果已列入表3, 从对照实际的情况看, 计算分类比原有分类较为合理, 而且达到了分类的数值化。此外, 在研究地区, 当需要进行新的土类鉴别时, 只要能取得各侵蚀因子的强度值, 就能利用上列方程组来判定其类型归属。例如, 一新的观测土壤, 其各因子强度记值如表7, 将各强度值分别代入方程组, 算得五个F值:

$$F_1 = 9.63, F_2 = 15.21, F_3 = 18.90, F_4 = 17.81,$$

表7 某一侵蚀土壤的侵蚀因子强度记值

侵蚀因子	土壤可蚀性	地面坡度	利用状况	干扰强度
强度记值	8	6	5	10

$F_5 = 13.54$ , 并称它们为相应土类的判别得分, 观测土壤应归属于得分最高的土类。此处得分最高者为18.90, 可见该土壤为中度侵蚀类型。

2. 土壤侵蚀趋向预报 不同侵蚀土壤类型是在各侵蚀因子相对稳定的条件下, 在一定时期内综合作用的结果。如果主要侵蚀因子或其相应强度没有明显的改变, 各侵蚀土壤类型在近期内也相对稳定, 因此, 某土类分布区的流失强度和流失量也与该土类一致。显然, 轻微侵蚀土类区的土壤流失强度和流失量最小。但是, 如果侵蚀因子强度产生了改变, 土壤流失强度和流失量也会改变, 其结果, 必然达到某一土类的流失水平。由于这样, 就可以利用判别方程根据各侵蚀因子强度来预报土壤侵蚀趋向。土壤侵蚀趋向有两个方面: 减轻或恶化。土壤侵蚀减轻是各侵蚀因子强度降低的结果; 土壤侵蚀恶化则是各侵蚀因子强度增高的结果。侵蚀趋向预报包括侵蚀因子强度变化后预报土壤流失量和流失强度将会与那一土类的相应量一致; 如果保持某一土类的流失强度和流失量的一定范围, 预报出各侵蚀因子强度允许改变的范围。例如, 在上例中该观测土壤, 地面坡度为10—15°, 记码值为6, 现若改变为0—5°, 记码值为1时, 通过计算, 说明土壤侵蚀强度和侵蚀量由此引起的变化, 并没有引起土类的改变(表8), 该土壤仍为中度侵蚀土壤。但是, 如果其它侵蚀因子强度指标不变, 利用干扰强度由强度级改变为轻度级, 这时, 相应的土壤侵蚀强度和侵蚀量则由中度改变至轻度侵蚀土类的相应量, 该土类也就改变为轻度侵蚀土类。判别得分值的改变见表8。相反, 对上述同一中度侵蚀土壤, 要使它的侵蚀强度和侵蚀量减到轻度侵蚀土类的水平, 只靠改变地面坡度由10—15°变为0—5°并不能达到目的, 而如果将利用干扰强度由强度级变为轻度级时, 即可满足要求。当然也可采用其它途径达到同一目的。

逐步判别分析只能对已有一定分类基础的研究对

表8 侵蚀因子强度指标改变后相应 $F_i$ 值的变化

$F_i$	$F_1$	$F_2$	$F_3$	$F_4$	$F_5$
按原有指标时	9.63	15.21	18.90	17.81	13.54
地面坡度记值由6变为1时	6.40	9.61	10.75	7.16	0.09
干扰强度记值由10变为1时	6.84	9.90	6.66	3.68	-5.81

(下转封3)

## 中国科学院环境科学委员会 第二次“农药与环境”学术报告会在宁召开

南京土壤研究所受中国科学院环境科学委员会的委托，主持筹备召开了第二次“农药与环境”学术报告会，会议于1982年10月27日至30日在南京举行，出席会议共17个单位45名代表。中国科学院南京分院院长、南京土壤研究所所长熊毅教授致开幕词，院环委会办公室主任卢贵钦同志出席了会议，会议邀请了江苏省农药所所长程煊生研究员，武汉医学院刘毓谷教授，武汉病毒所副所长简浩然研究员，上海昆虫所副所长刘维德副研究员在会上作国内外有关农药与环境领域的

研究动态报告。会议收到论文39篇，其中18篇在全体会议上进行了学术交流，论述了区域环境中农药残留、迁移和消失动态，农药的降解与代谢，农药的生态效应（包括对水生生物，动物，微生物及其酶），农药厂的污水处理及农药残留监测方法等研究。最后，针对我国1985年将禁止使用六六六，滴滴涕农药的决定，与会者对农药与环境研究工作的方向和任务进行了热烈的讨论。

（张水铭）

（上接第109页）

象进行类型判定和类型调整。同时，所建立的判别方程组总是包含多个方程，而且随所要判定的类型数目的增加而增加，因而使得进行类型判别计算很繁复。这一问题用电子计算机通过计算编制出类型检索表而得到了解决〔6〕，利用该表，使类型判定无需经过计算，直接由查表得知。此外，本文还涉及数码如何取得和离散数码值使之连续化以提高精度问题，这些问题与文字资料的数值化和离散值的进一步处理有关，已有另文专门讨论。

利用逐步判别分析，以各侵蚀因子强度为指标，可使侵蚀土壤分类数值化。通过所建立的判别方程，可以对野外的土类划分进行调整、判定新取土壤的类型归属和预报侵蚀土壤的发展趋势。因此这一分析方法对进行土壤资料处理，对水土保持工作有重要价值。作

为一种方法还能在资源评价、土地评级和土壤发生分类等工作中应用。

### 参考文献

- 〔1〕中国科学院地质研究所编著：数学地质引论，128—156页，地质出版社，1977。
- 〔2〕于崇文等编著：数学地质的方法与应用，115—135页，冶金工业出版社，1980。
- 〔3〕R. Webster: Quantitative and numerical methods in soil classification and survey, Clarendon Press Oxford, 1977.
- 〔4〕J. H. Rayner: J. Soil Sci., 17: 77—92, 1966.
- 〔5〕杨艳生等：关于土壤流失方程中K因子的探讨。中国水土保持，第4期，39—42，1982。
- 〔6〕杨艳生等：宁夏固原县侵蚀土壤的数值分析研究。土壤学报，19(1): 71—84, 1982。